Spatial Statistics 9 (2014) 122–145

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Spatial econometric panel data model specification: A Bayesian approach

James P. LeSage

Fields Endowed Chair, Texas State University—San Marcos, Department of Finance & Economics, San Marcos, TX 78666, United States

ARTICLE INFO

Article history:
Received 21 August 2013
Accepted 26 February 2014
Available online 6 March 2014

Keywords:
Static space–time panel data models
Bayes factors
Local versus global spatial spillovers

ABSTRACT

Taking a Bayesian perspective on model uncertainty for static panel data models proposed in the spatial econometrics literature considerably simplifies the task of selecting an appropriate model. A wide variety of alternative specifications that include various combinations of spatial dependence in lagged values of the dependent variable, spatial lags of the explanatory variables, as well as dependence in the model disturbances have been the focus of a literature on various statistical tests for distinguishing between these numerous specifications.

A Bayesian model uncertainty argument is advanced that logically implies we can simplify this task by focusing on only two model specifications. One of these, labeled the spatial Durbin model (SDM) implies *global spatial spillovers*, while the second, labeled a spatial Durbin error model (SDEM) leads to *local spatial spillovers*. A Bayesian approach to determining an appropriate local or global specification, SDEM versus SDM is set forth here for static panel variants of these two models. The logic of the Bayesian view of model uncertainty suggests these are the only two specifications that need to be considered. This greatly simplifies the task confronting practitioners when using static panel data models.

E-mail addresses: james.lesage@txstate.edu, jlesage@spatial-econometrics.com.

http://dx.doi.org/10.1016/j.spasta.2014.02.002

1. Introduction

A Bayesian model uncertainty argument is advanced that logically implies we can simplify the task of model specification by focusing on only two model specifications. One of these, labeled the spatial Durbin model (SDM) implies *global spatial spillovers*, while the second, labeled a spatial Durbin error model (SDEM) leads to *local spatial spillovers*. A spatial spillover arises when a causal relationship between characteristics/actions of entity/agent (X_i) located at position i in space exerts a significant influence on the outcomes/decisions/actions (Y_j) of an agent/entity located at position j. Formally, LeSage and Pace (2009) define this as: $\partial Y_j/\partial X_i \neq 0$. If locations j are neighbors to location i, we have a *local* spatial spillover. In contrast, if locations j include not only neighbors to i, but neighbors to neighbors of i, neighbors to neighbors to neighbors, and so on, we have a *global* spillover. The logic of the Bayesian view of model uncertainty suggests these are the only two specifications that need to be considered. This greatly simplifies the task confronting practitioners when using static panel data models.

A Bayesian approach to determining an appropriate local or global specification, SDEM versus SDM is set forth here for static panel variants of these two models. The approach taken is to rely on analytical integration of regression parameter β, and noise variance parameter σ, in the marginal likelihood used to calculate posterior model probabilities. A final step involves univariate numerical integration over a scalar spatial dependence parameter. This approach differs from common practice where numerous alternative (approximate) approaches have been proposed for calculating model probabilities. These alternatives are popular because of general pessimism about the ability of numerical integration to produce accurate results for model comparison (see Kass and Raftery, 1995). A series of experiments is carried out to investigate the numerical accuracy of the approach set forth here. The results indicate that given a sufficient level of spatial dependence, the approach produces accurate conclusions regarding the appropriate specification.

Two of the most popular spatial regression model specifications are: (1) those involving spatial lags of the dependent variable, and (2) those involving spatial lags of the disturbances. The theoretical motivation and substantive consequences of these two competing specifications are quite different. A model such as the spatial autoregressive (SAR) specification in (1) that includes a spatial lag of the dependent variables implies what has come to be labeled *endogenous interaction* effects.

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + X_{it}\beta + \nu_i + \psi_t + \varepsilon_{it} \tag{1}$$

$$y_{it} = X_{it}\beta + \mu_i + \lambda_t + u_{it}, \quad u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it} \tag{2}$$

$$\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2 I_{NT}),$$

where w_{ij} is the $(i, j)th$ element of the spatial weight matrix reflecting spatial proximity of the N regions that make up the panel of regions over $t = 1, \ldots, T$ time periods. The diagonal elements $w_{ii} = 0$, and the matrix is normalized to have row-sums of unity. The parameters $\nu_i, i = 1, \ldots, N$ are fixed effects for the regions and $\psi_t, t = 1, \ldots, T$ time-period specific effects. Interaction effects are reflected in the spatial lag variable $\sum_{j=1}^{N} w_{ij} y_{jt}$ (and associated scalar parameter ρ), which allows dependent variable values of spatial unit i to be influenced by those of spatial units $j = 1, \ldots, n$.

Expression (2) shows a spatial error specification (SEM) where there are no spatial spillovers, but rather interaction effects involving the disturbances, with the strength of this type of dependence reflected in the scalar parameter λ. This specification allows for shocks (disturbances) of region $j = 1, \ldots, N$ to influence the disturbances of region i. Given the LeSage and Pace (2009) definition of spatial spillovers: $\partial y_j/\partial X_i \neq 0$, it should be clear that the SEM model sets cross-partial derivatives such as this to zero, a consequence of the independence assumption made by regression models.

To see how the SAR specification produces non-zero cross-partial derivatives, we need to consider the reduced form. Using matrix notation, the SAR specification from (1) can be written as in (3), with the partial derivative for the kth explanatory variable shown in (5). As noted by Elhorst (2012), the